UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/634,145	08/04/2003	Chew Kiat Heng	NAA 0018 PA/41049.20	5097

23368 7590 05/11/2010
DINSMORE & SHOHL LLP
FIFTH THIRD CENTER, ONE SOUTH MAIN STREET
SUITE 1300
DAYTON, OH 45402-2023

EXAMINER	
WHALEY, PABLO S	

ART UNIT	PAPER NUMBER
1631	

MAIL DATE	DELIVERY MODE
05/11/2010	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/634,145	HENG ET AL.	
	Examiner	Art Unit	
	PABLO WHALEY	1631	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on _27 January 2010_.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) _1-15,17-26 and 28-30_ is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) _1-15, 17-26, and 28-30_ is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

Status of Claims

Claims 1-15, 17-26, and 28-30 are pending and under consideration.

Claims 16 and 27 are cancelled.

Withdrawn Rejections

The rejection of claims 1-15, 17-26, and 28-30 under 35 U.S.C. 112, second paragraph, is withdrawn in view of applicant's amendments filed 01/27/2010.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.
>
> This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g) prior art under 35 U.S.C. 103(a).

The factual inquiries set forth in *Graham* **v.** *John Deere Co.*, 383 U.S. 1, 148 USPQ 459 (1966), that are applied for establishing a background for determining obviousness under 35 U.S.C. 103(a) are summarized as follows:

1. Determining the scope and contents of the prior art.
2. Ascertaining the differences between the prior art and the claims at issue.

3. Resolving the level of ordinary skill in the pertinent art.
4. Considering objective evidence present in the application indicating obviousness or nonobviousness.

Claims 1-6, 9-11, 13-15, 17-21, and 28-30 are rejected under 35 U.S.C. 103(a) as being made obvious by Parzen (Biometrics, 1999, Vol. 55, p.580-584), in view of Shattuck-Eidens et al. (JAMA, 1997, Vol. 278, No. 15, p. 1242-1250), and in view of Cleveland (Journal of the American Statistical Association, 1979, Vol. 74, No. 368, p.829-836).

The instantly rejected claims are drawn to a computer-implemented method of determining a statistical model for predicting disease risk for a member of a population. The claims require collecting non-genetic data, genetic data, and data that indicates disease status. The claims require storing a candidate statistical model depending on a plurality of parameters for calculating disease risk as a function of non-genetic data. The claims require optimizing model parameters by fitting, where the fitting is based on calculating a deviate of a predicted risk from an indicator of disease status for each set by using the candidate model and non-genetic data; calculating a sum of weighted deviates for all sets, where the weights are associated with the set for which each deviate has been calculated; and determining weights used to weight the deviates with a constraint such that sets with the same genetic data have the same weights. Additionally, the claims require that optimum parameters are obtained by minimizing the sum of weighted deviates and used with the candidate model for calculating disease risk.

Parzen teaches a method for optimizing linear regression models used to predict liver disease [Abstract]. In particular, Parzen shows a Cox hazard regression model for calculating disease risk in a subject are described in full [Section 2]. The model is a linear combination of coefficients and covariates that include age, albumin, and edema data sets [Table 1, Table 2, and p.581, Col. 2], which are interpreted as non-genetic risk factors. An indicator of disease status is described, N(t), which fluctuates between 1

and 0 over time based on patient risk [p.581, Col. 2]. Parzen calculates partial risk estimates using a Cox likelihood score vector wherein Z is based on a sum of weighted averages and dN is a binary variable between 1 and 0 (i.e. weight) [p.581, Col. 2, Equation 2], which is interpreted as a target function. Parzen describes an optimization procedure based on curve-fitting [Section 3]. In particular, data is partitioned into groups and group weights (I) are assigned a value of 1 or 0 [p.581, last ¶]. If the model is correctly specified, parameters for an arbitrary number of groups will take the value of zero in the Cox model [p.582, Col. 1, ¶1 and Equation 3], which shows weights associated with sets of data having like values. Subjects in the same group can also be considered similar if they have similar risks at any given time [p.581, Col. 1]. Parzen also calculates the Chi-squared distribution as an alternative measure of goodness of fit [p.582, Col. 1]. Parzen also defines a residual equation for calculating goodness of fit based the difference between observed minus expected number of failures in each region [p.582, Col. 2], and calculates the total number of failures based on the sum of the estimated expected failures. The Chi-squared distribution is interpreted as a teaching for calculating weighted deviates since it is used in the model fitting process and is based on weighted deviations in the data. Parzen shows selecting the model with a minimized goodness of fit statistic [p. 582, Col. 1 and 583, Col. 1, ¶2]. Parzen shows two different models with the same number of parameters [Equations 1 and 3].

Parzen does not teach collecting genetic data sets associated with members of a population, as in claims 1, 17, 21, and 28.

Parzen does not teach a fitting procedure that includes determining weights used to weight the deviates with a constraint such that the sets that have the same genetic data have the same weights, as in claims 1, 21, and 28.

Parzen does not teach a fitting procedure wherein weights are weighted by an adjustment factor, as in claims 13 and 14.

Shattuck-Eidens teaches collecting genetic data related to disease [See at least Table 1] and a statistical model for predicting disease risk that accounts for the incidence of groups with different types of genetic mutations and non-genetic factors [p.1243, Col. 3 and p.1244, Col. 1 and 2]. In particular, the incidence of genetic disorders is represented using integer values [p.1246, Col. 2 and 3], which are interpreted as data associated with genetic factors in light of the specification [0034]. Additionally, groups with the same genetic disorder are assigned a similar integer value [p.1246, Col. 2 and 3], which reasonably suggest a constraint such that sets with the same genetic data have the same weights. This method is beneficial for predicting cancer in patients with detected genetic mutations [p.1244].

Cleveland teaches a computer-based method for optimizing models based on weighted regression. Cleveland shows weight functions represented as integers [p.829] and shows calculating a sum of weighted deviates [p.830, Col. 2]. A statistical model is optimized by re-fitting the regression model using the newly calculated weighted deviate values [See steps 1-4, p. 830 and 831]. Cleveland also shows a weighting function wherein values above a certain x threshold all equal 0 [p.831, Col. 1], which suggests equal weights for certain points in a data set. Cleveland also shows an optimization process that includes a robustness weight calculation that is used to weight different weights and is based on a ratio of residuals and the median [p.831, Col. 1], which shows weights weighted by an adjustment factor. Cleveland further optimizes parameters based on error variance and linear sum of residuals [Section 6.1 and p.835, Col. 1]. This technique is beneficial for smoothing distortions in data [Section 4.4]. Cleveland shows techniques for reducing computations [Section 5.1], which inherently shows the use of computers and computer software for performing these methods.

It would have been obvious to someone of ordinary skill in the art at the time of the instant invention to collect genetic data sets associated with members of a population, as taught by Shattuck-Eidens [See at least Table 1], in the method taught by Parzen, since Shattuck-Eidens shows such data can be used in a statistical model that accounts for the incidence of groups with different types of genetic

mutations and non-genetic factors with predictable results [p.1243, Col. 3 and p.1244, Col. 1 and 2]. The motivation would have been to use regression models for describing the relationship between multiple variables related to disease, as suggested by Shattuck-Eidens [p.1246].

It would have been obvious to someone of ordinary skill in the art at the time of the instant invention to determine weights with the constraint such that the sets that have the same genetic data have the same weights in the method taught by Parzen, since both Parzen [p.581, Col. 2] and Cleveland [p. 830-831] show that group weights in weighted deviate calculations can take on equal values and are subject to data-dependent constraints, and since Shattuck-Eidens teaches groups with the same genetic disorder assigned a similar integer value with predictable results [p.1246, Col. 2 and 3], which reasonably suggests genetic populations with equal weights. The motivation would have been to improve the disease model by finding values for the coefficients such that the regression model matches the raw data as well as possible, as suggested by Cleveland [p.830].

It would have been obvious to someone of ordinary skill in the art at the time of the instant invention to calculate new weights that are weighted by an adjustment factor, as taught by Cleveland [p.830, Col. 1], in the method of Parzen, where the motivation would have been to improve model performance through robust local regression, as suggested by Cleveland [p.830].

It would have been obvious to someone of ordinary skill in the art at the time of the instant invention to practice the method made obvious by Parzen, Shattuck-Eidens, and Cleveland using a computer and computer software since Shattuck-Eidens and Cleveland suggests such predictive methods are designed for computers. The motivation would have been to improve disease prediction using automated techniques for performing complex calculations.

Claims 7, 8, 12, and 22-26 are rejected under 35 U.S.C. 103(a) as being made obvious by Parzen (Biometrics, 1999, Vol. 55, p.580-584), in view of Shattuck-Eidens et al. (JAMA, 1997, Vol. 278, No. 15, p. 1242-1250), and in view of Cleveland (Journal of the American Statistical Association, 1979, Vol. 74, No. 368, p.829-836), as applied to claims 1-6, 9-11, 13-15, 17-21, and 28-30, above, and further in view of Kooperberg et al. (Technical Report, 1996, p.1-20) and Hu et al. (Proceedings of the Survey Research Methods Section, ASA, 1996, p.287-292).

Parzen, Shattuck-Eidens, and Cleveland make obvious a method for determining a model for predicting disease risk, as set forth above. Additionally, Cleveland shows a function based on a summation of weights and residual size [See at least p.830, Col. 1, Col. 2, and p.834, Col. 1]. Shattuck-Eidens also shows correlating risk factors and grouping risk factors based on clustering [p.1246 and Table 6].

Parzen, Shattuck-Eidens, and Cleveland do not teach a residual for an i-th one of said data sets in said reference group that is the difference between a value of the indicator of disease status contained in said i-th data set and the value of disease risk for the member associated with said i-th data set, where the value of disease risk is calculated from said candidate model with parameters optimized for a given set of group weights by fitting data sets in groups other than the reference group, as in claim 7.

Parzen, Shattuck-Eidens, and Cleveland do not teach imputing missing data, as in claims 12 and 22.

Parzen, Shattuck-Eidens, and Cleveland do not teach dividing data, and recursive division, as in claims 23, 24, 25, and 26.

Parzen, Shattuck-Eidens, and Cleveland do not teach determining if a criterion is met after dividing, said criterion evaluated based on genetic data in each of said data sets, and regrouping when criteria are not met, as in claim 23.

Parzen, Shattuck-Eidens, and Cleveland do not teach performing division recursively on each group of a division, and wherein divisions are made dependent on data indicative of different factors, as in claims 24-26.

Methods for dividing data within predictive modeling processes are well known. In particular, Kooperberg teaches methods for selecting optimal models by dividing data into equally sized subgroups with the constraint that data not in the j-th subgroup is fitted to the model [See Section 3.2, p.6-7], which is interpreted as fitting data sets in groups other than the reference group. The best model is selected by minimizing a cross-validation loss function for data not used to fit the model [See Section 3.2, p.6-7].

Methods for imputing data with predictive modeling processes are well known. In particular, Hu shows software for imputing missing values in regression models. The software program partitions the range of regression values from the data set into subsets [p.287, Col. 2, ¶2, p.288, Col. 2, ¶2]. Weighted average values are then computed and assigned to subsets with missing data [p.287, Col. 2, ¶2]. The subsets are assumed to be homogeneous. This technique is beneficial for eliminating bias in large data sets [Section II and p.292, Col. 2].

It would have been obvious to someone of ordinary skill in the art at the time of the instant invention to determine a residual for an i-th data sets in a reference group that is the difference between a value of the indicator of disease status contained in said i-th data set and the value of disease risk for the member associated with said i-th data set, as suggested by Parzen [p.582, Col. 2] and Cleveland [p.830], where the value of disease risk is calculated from said candidate model with parameters optimized for a given set of group weights by fitting data sets in groups other than the reference group, as suggested by

Kooperberg [See Section 3.2, p.6-7], in the method made obvious by Parzen, Shattuck-Eidens, and Cleveland, where the motivation would have been to employ cross-validation as an alternative risk estimation, as suggested by Kooperberg [See Section 3.2, p.6-7].

It would have been obvious to someone of ordinary skill in the art at the time of the instant invention to imput missing data in regression models, as taught by Hu, in the method made obvious by Parzen, Shattuck-Eidens, Cleveland, and Kooperberg, since Shattuck-Eidens uses data that includes missing data sets with predictable results [Table 5]. The motivation would have been to eliminate bias in large data sets [Section II and p.292, Col. 2].

It would have been obvious to someone of ordinary skill in the art at the time of the instant invention to divide data based on a criteria, as taught by Cleveland [Section 5.1] and Kooperberg [Section 3.2], in the method made obvious by Parzen, Shattuck-Eidens, Cleveland, and Kooperberg, where the motivation would have been to reduce the computational load on a computer, as suggested by Cleveland [Section 5.1].

Response to Arguments

Applicant's arguments filed 01/27/201 have been fully considered but are not persuasive for the following reasons.

In response to applicant's statement that Shattuck-Eidens does not teach genetic factors [p.11] and applicant's reference to the specification [0035], it is noted that the claimed statistical model comprises non-genetic data and does not specifically use genetic factors [See claim 1]. Furthermore, the specification describes genetic factors as data entries represented as integers or other data formats [0034]. Shattuck-Eidens teaches the collection of genetic data related to disease [See at least Table 1] and a statistical model for predicting disease risk that accounts for the incidence of groups with different types of genetic mutations and non-genetic factors [p.1243, Col. 3 and p.1244, Col. 1 and 2]. In particular, the

incidence of genetic disorders is represented using integer values [p.1246, Col. 2 and 3], which meets the claim language for genetic factors as interpreted in light of the specification. Furthermore, these integer values are associated with patients and are used in the statistical model [p.1246, Col. 2 and Col. 3].

In response to applicant's statement [p.12] that none of the previous Office actions have pointed to any teachings in the cited references that show predicting risk for a member of the population using a model and non-genetic data associated with that member, Parzen teaches a regression model for predicting disease in individuals based on non-genetic risk factors, as set forth above. It is noted that the claimed statistical model does not specifically use genetic data. Additionally, Shattuck-Eidens teaches the collection of genetic data as well as a statistical model that uses both genetic factors and non-genetic factors, as discussed above.

For these reasons, the examiner maintains that the combination of references teaches and/or makes obvious the claimed limitations.

Conclusion

No claims are allowed.

THIS ACTION IS MADE FINAL. Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX

MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Pablo Whaley whose telephone number is (571)272-4425. The examiner can normally be reached between 12pm-8pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Marjorie Moran can be reached at 571-272-0720. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Pablo S. Whaley

Patent Examiner

Art Unit 1631

/PW/

/Marjorie Moran/
Supervisory Patent Examiner, Art Unit 1631